W338–W344 *Nucleic Acids Research, 2019, Vol. 47, Web Server issue* *Published online 22 May 2019*
doi: 10.1093/nar/gkz383

mCSM-PPI2: predicting the effects of mutations on protein–protein interactions

Carlos H.M. Rodrigues[1,2,3], Yoochan Myung[1,2,3], Douglas E.V. Pires[1,2,3,*] and David B. Ascher[1,2,3,4,*]

[1]Department of Biochemistry and Molecular Biology, University of Melbourne, Melbourne, Australia, [2]ACRF Facility for Innovative Cancer Drug Discovery, Bio21 Institute, University of Melbourne, Melbourne, Australia, [3]Structural Biology and Bioinformatics, Baker Heart and Diabetes Institute, Melbourne, Australia and [4]Department of Biochemistry, University of Cambridge, Cambridge, UK

Received February 11, 2019; Revised April 30, 2019; Editorial Decision April 30, 2019; Accepted May 20, 2019

ABSTRACT

Protein–protein Interactions are involved in most fundamental biological processes, with disease causing mutations enriched at their interfaces. Here we present mCSM-PPI2, a novel machine learning computational tool designed to more accurately predict the effects of missense mutations on protein–protein interaction binding affinity. mCSM-PPI2 uses graph-based structural signatures to model effects of variations on the inter-residue interaction network, evolutionary information, complex network metrics and energetic terms to generate an optimised predictor. We demonstrate that our method outperforms previous methods, ranking first among 26 others on CAPRI blind tests. mCSM-PPI2 is freely available as a user friendly webserver at http://biosig.unimelb.edu.au/mcsm_ppi2/.

INTRODUCTION

Most biological processes, including cell proliferation (1), signalling (2), host-pathogen interactions (3) and protein transport (4), are intrinsically coordinated through complex networks of protein–protein interactions. The diversity and size of the interactome offers a highly selective and tunable way to modulate protein activities and pathways (5). Genetic variations leading to changes in the binding affinity of these interactions can disrupt or directly affect the formation of interacting complexes and consequently lead to disease (6–16) and drug resistance (17–19).

Advances in next-generation sequencing techniques have created an explosive increase in the number of genetic variants available in the literature. However, experimental techniques to study these variants are still expensive and time consuming. mCSM (20) was one of the first scalable computational tools to accurately predict the effects of mutations on binding affinity. Previous methods were limited either in terms of their throughput (21,22) or in terms of their performance (23). Since then, significant efforts have been devoted to computationally study the effects of mutations on protein complexes (24,25) but their poor predictive performance on new variants, particularly mutations that lead to increased binding affinity of the complex, has limited their use. In addition, the increase in amount of experimental evidence of effects of variants on binding affinity offers the opportunity to develop new and more accurate methods.

Our previously described graph-based signatures concept has proven to be a powerful approach and has been widely applied to the study of protein structure, including how mutations alter protein stability (20,26), dynamics (27) and interactions with other molecules (20,28–34).

Here we introduce mCSM-PPI2, a webserver that integrates our well-established mCSM graph-based based signatures framework with evolutionary information, inter-residue non-covalent interaction networks analysis and energetic terms, in order to provide an optimized overall prediction performance.

MATERIALS AND METHODS

Data sets

The data used on this work was derived from the recently updated version of the SKEMPI database (35), which compiles experimental data on changes in thermodynamic and kinetic parameters on mutation for protein–protein complexes that have 3D structures deposited in the PDB. SKEMPI 2.0 (36) includes new mutations identified in the literature after its first release, including data available from three other databases: ABbind (37), PROXiMATE (38) and dbMPIKT (39). The average mutation effect was considered for variants reported in multiple experiments when these varied by less than 2.0 kcal/mol and discarded otherwise. After filtering for only single-point mutations with available

*To whom correspondence should be addressed. Tel: +61 90354794; Email: david.ascher@unimelb.edu.au
Correspondence may also be addressed to Douglas E.V. Pires. Email: douglas.pires@unimelb.edu.au

experimental crystal structures of the wild-type, we were able to collect 4169 (S4169) variants in 319 different complexes. All protein structures were collected from the Protein Data Bank and a series of pre-processing steps were performed to account for the diversity of structures (see Supplementary material).

The binding affinity of protein–protein complexes were used to calculate the binding Gibbs free energy (ΔG):

$$\Delta G = RT \ln (K_D)$$

where $R = 1.985 \times 10^{-3}$ kcal $K^{-1}mol^{-1}$ is the ideal gas constant, T is the temperature (in K) and K_D is the equilibrium dissociation constant of the protein–protein complex (in molar). The change in binding affinity upon mutation was calculated as follows:

$$\Delta\Delta G_{wt-mt} = \Delta G_{wild-type} - \Delta G_{mutant}$$

Since the Gibbs free energy formulation is a thermodynamic state function a change in binding affinity of a mutation from a wild-type protein to its mutant ($\Delta\Delta G_{WT \rightarrow MT}$) should be equivalent to the negative change binding free energy of the hypothetical reverse mutation, from the mutant to the wild-type protein ($-\Delta\Delta G_{MT \rightarrow WT}$) (40). Given the unbalanced nature of the original dataset collected from SKEMPI 2.0, 901 variants increased ($\Delta\Delta G_{wt\text{-}mt} \geq 0$) and 3268 decreased ($\Delta\Delta G_{wt\text{-}mt} < 0$) binding affinity, and in order to build a more robust and balanced predictive method, we also included the hypothetical reverse mutations. Therefore, the final dataset for building mCSM-PPI2 predictive model includes 8338 single-point mutations (S8338), which represents an increase of up to three-fold in datapoints in comparison with previous methods that used data from the first version of SKEMPI with 2007 (S2007) (20,23), 1964 (S1964) (25), 1102 (S1102) (24) and 1327 (S1327) mutations (41).

A subset of 487 mutations in 56 complexes (S487) contained within S4169 and not in S2007 were recently curated (24) and here we used as evidence to evaluate the performance of mCSM-PPI2. A summary of different subsets derived from SKEMPI is shown in Supplementary Table S1.

The datasets used in this work are freely available for download at http://biosig.unimelb.edu.au/mcsm_ppi2/datasets.

Graph-based structural signatures

mCSM-PPI2 uses as one of its core components our well established graph-based structural signatures (mCSM) to represent the environment of the wild-type residue. This approach models both the geometry and physicochemical properties of the interactions and architecture of wild-type structure and has been widely applied to the study of small molecule and protein structure (20,26–34,42). Our signatures represent atoms as nodes and their interactions as edges, with their physicochemical properties encoded based upon the amino acid residue properties, denoted by a pharmacophore. From this representation of the residue environment, distance patterns between atoms characterized by their properties are summarized in concise signatures as cumulative distribution functions.

Modelling effects of mutation

Single-point mutations can affect protein–protein interactions via different molecular mechanisms, including changing folding free energy of interacting partners or disrupting non-covalent interactions essential for complex formation (6,43). In mCSM-PPI2, we have included six new distinct types of features that were not used in our first method (Supplementary Table S2). These were combined with our well-established graph-based signatures as evidence for training a machine learning algorithm (see Supplementary material) to better explore the effects of mutations in protein–protein binding affinity (Figure 1).

Wild-type residue environment. Based on 3D structures collected from the Protein Data Bank (44), we were able to calculate Relative Solvent Accessibility (RSA), torsion angle PHI and residue depth for the wild-type residue using BioPython (45) version 1.7. We also extracted information on the amino acid content in the sequence of the chain in which the mutation occurs using iFeature (46).

Nature of wild-type and mutant residues. The conformational flexibility of glycine side chains and the rigidity of proline side chains are important for defining the backbone flexibility. Mutations from and to these two amino acids can lead to large structural effects. For our model we included binary terms to capture if the mutation was from or to a glycine or proline.

Evolutionary information. Binding regions are known to be evolutionarily conserved, which has been exploited in a variety of studies to identify potential protein interaction interfaces. For mCSM-PPI2 we also harnessed this information by using the Position Specific Scoring Matrix (PSSM) scores. PSSM was calculated through PSI-BLAST of BLAST 2.2.3 using the non-redundant Swiss-Prot database of protein sequences and the sequence of the chain in which the mutation occurs as the query parameter.

Non-covalent interaction network analysis. We performed analysis of the non-covalent interactions for the wild-type residue and for the closest interface using the contacts calculated by Arpeggio (47). Here, we extracted two types of information: the difference between the number of contacts of wild-type and mutant residue for covalent, Van der Waals', aromatic and hydrogen bond contacts, and complex network metrics for the contact graph of the closest interface of interaction, from which we extracted centrality metrics, including closeness and central points (48). In this work, we consider a residue to be at an interaction interface if it is located at most 5 Å away from the interacting partner, following previous studies. In addition, we included three protein contact potentials scores from the AAindex database (49) (Supplementary Table S3).

Energetic terms. Interaction energy information between the two interacting chains were extracted from FoldX (22). In addition, we included the predicted folding free energy change upon mutation.

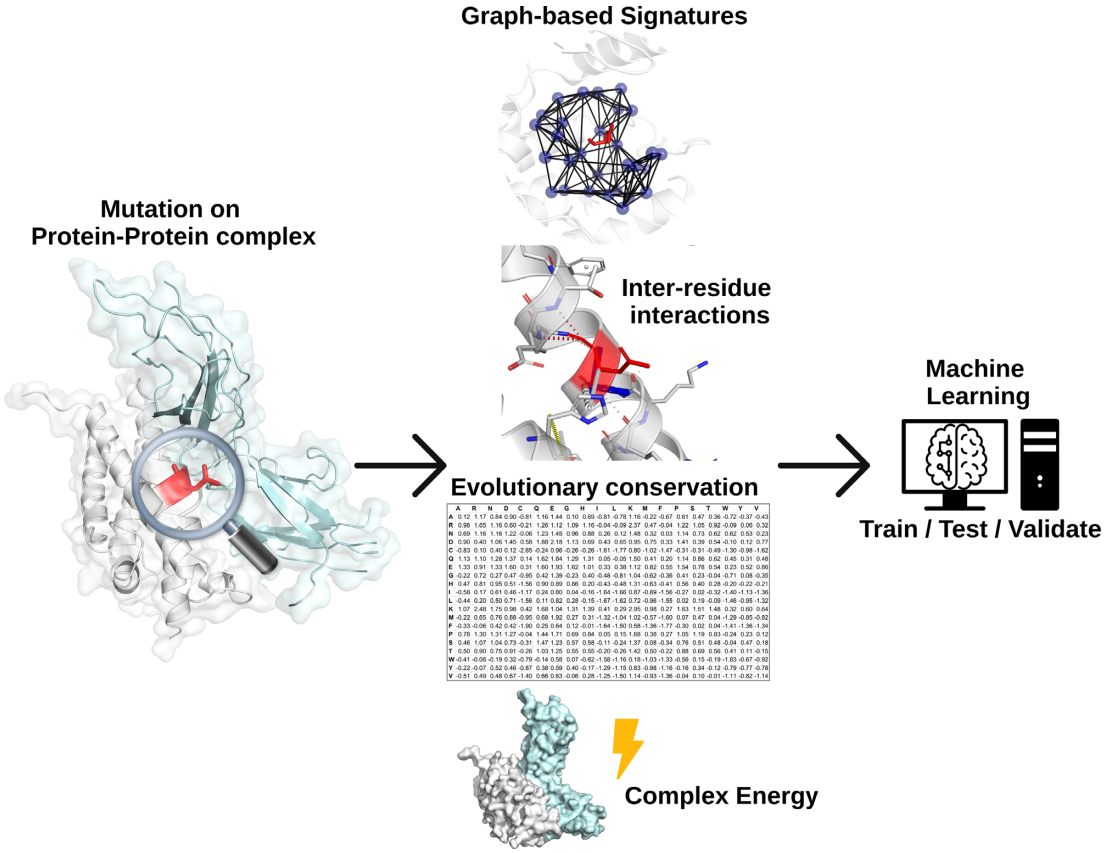

Figure 1. mCSM-PPI2 methodology workflow. The method relies on graph-based signatures, which model distance patterns and encode geometrical and physico-chemical properties on wild-type residue environment. Network analysis based on non-covalent interactions of wild-type residue and interacting interface along with evolutionary information and energy terms are also used. All features are used as evidence to train, test and validate machine learning algorithms.

Atomic fluctuation. We used the Bio3D R package (50) to calculate the atomic fluctuations of the structure of the monomer where the mutation occur using calpha and pfanm force fields to account for effects on protein flexibility/rigidity.

WEBSERVER

We have implemented mCSM-PPI2 as a user-friendly and freely available webserver (http://biosig.unimelb.edu.au/mcsm_ppi2/). The server front end was built using Materialize framework version 1.0.0, while the back-end was built in Python via the Flask framework (Version 1.0.2). It is hosted on a Linux server running Apache.

Input

mCSM-PPI2 can be used in two different ways: to either assess the effects of mutations specified by the user input or to predict the effects of mutations at the protein–protein interface in an automated manner. For user-specified variations two options are available (Supplementary Figure S1). The 'Single Mutation' option requires one to provide a PDB file or PDB accession code of the structure of the protein complex, the point mutation specified as a string containing the wild-type residue one-letter code, its corresponding residue number and the mutant residue one-letter code. The 'Mutation List' option allows users to upload a list of mutations in a plain text file for batch processing. For both options, users are also required to specify the chain identifier in which the wild-type residues are located.

Alternatively, for assessing effects of mutations at protein–protein interfaces the server requires the user to provide a PDB file or PDB accession code and select one of two options: alanine scanning (all interface residues are mutated to an Alanine) or saturation mutagenesis (all interface residues are mutated to every other amino acid) (Supplementary Figure S2).

In order to assist users to submit their jobs for predictions, sample submission entries are available in both submission pages and a help page is also available via the top navigation bar.

Output

For the Single Mutation option (Supplementary Figure S3), mCSM-PPI2 outputs the predicted change in binding affinity (in kcal/mol) along with an interactive 3D viewer, built using NGL viewer (51), showing non-covalent interactions, generated with Arpeggio, at the mutated position. A set of controllers are available for users to hide and show the different types of interactions and to alternate between wild-

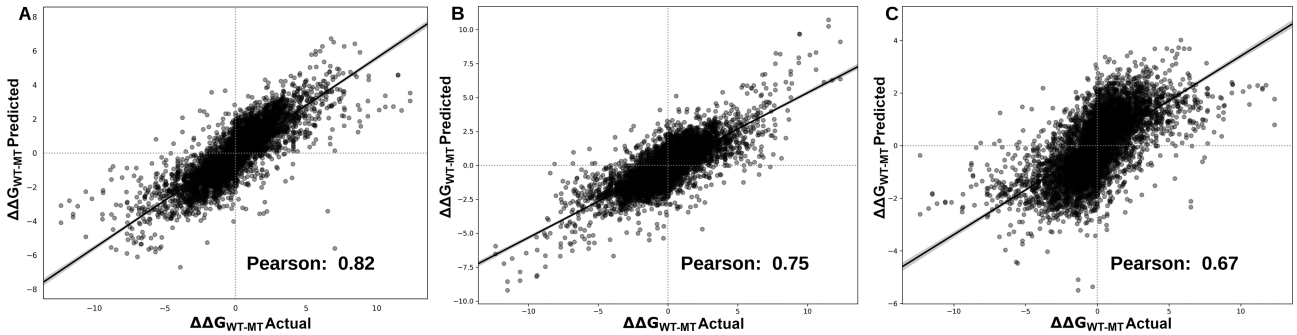

Figure 2. Performance evaluation on cross-validation. mCSM-PPI2 was able to achieve a Pearson's correlation of 0.82 and RMSE 1.18 kcal/mol when trained on the S8338 dataset applying 10-fold cross-validation 10 times (**A**). In low-redundancy sets, mCSM-PPI2 was able to achieve a correlation of 0.75 and 0.67 on leave-one-complex-out (**B**) and leave-one-binding-site-out (**C**), respectively.

type and mutant structures. In addition, a 2D viewer displaying non-covalent interactions of wild-type and mutant structures is also shown. Pymol sessions are available for download. For the Mutation List option (Supplementary Figure S4), the results are summarized in a downloadable table from which users can access details for each single variant.

For the Alanine Scanning option on the interface analysis, the server first presents a table with all the interfaces identified on the submitted structure, and it also allows for inspection of the individual interfaces. On the results page of each interface the server shows a downloadable table with the prediction outcomes for each mutation, a bar chart that summarizes the predicted changes in binding affinity (Supplementary Figure S5) and an interactive 3D viewer in which the residues are coloured according to the predicted value (Supplementary Figure S6). Similarly, for the Saturation Mutagenesis option, mCSM-PPI2 outputs a table with all the interfaces identified and allows the users to access detailed information on each interface. For each interface, the server outputs a table compiling the results for all variants (Supplementary Figure S7), a heatmap of all interface residues and their respective mutations (Supplementary Figure S8), and a 3D viewer in which the residues are coloured according to the average prediction for each particular position (Supplementary Figure S9).

VALIDATION

Performance on cross-validation

In order to build a more robust and reliable predictive model we performed four types of validation. Firstly, we performed 10-times stratified 10-fold cross-validation, using 90% of our original dataset (S8338) for training and the remaining as a blind test. Selection of the blind test was repeated 10 times in a stratified manner, with the model retrained on the remaining data, in order to test the robustness of the model (see Supplementary material). For this approach the hypothetical reverse mutations were kept in either training or test sets during the splits according to its counterpart original mutation. Our method was able to achieve an average Pearson correlation (ρ) of 0.82 with a standard deviation (σ) of 0.06 across the 10 runs (Figure 2A) showing a more balanced prediction when distinguishing between mutations that increase binding affinity from decreasing ones than other methods (Supplementary Table S4). We also evaluate the performance of mCSM-PPI2 when trained only on the original subset of mutations from SKEMPI2 (S4169) using the same procedure and obtained a correlation of 0.76 and RMSE of 1.19 kcal/mol. These results corroborate the use of reverse mutations in order to improve performance and robustness of our predictive model. Performance comparison between mCSM-PPI2 and other methods on different versions of SKEMPI and performance of individual types of attributes are shown in Supplementary Tables S5 and S6, respectively.

We further evaluated the performance of our approach on two low-redundancy sets; low redundant at the (i) complex and (ii) interface level. The complex low redundancy test was performed using leave-one-complex out cross-validation, in which we trained our model on 318 complexes of our dataset and evaluate the performance on the one remaining complex. After repeating this procedure for each complex we achieved $\rho = 0.75$ (Figure 2B) and Root Mean Square Error (RMSE) of 1.30 kcal/mol, outperforming MutaBind (25) ($\rho = 0.68$ and RMSE = 1.57 kcal/mol).

Similarly, we applied leave-one binding site out using the 'hold-out' information extracted from SKEMPI2. Here, we removed all mutations located in identical binding sites for testing and trained on the remaining data. mCSM-PPI2 achieved $\rho = 0.67$ (RMSE = 1.39 kcal/mol) (Figure 2C), which was significantly higher (p-value < 0.0001 by Fisher *r*-to-*z* transformation) than the results reported for MutaBind when trained using only mutations from SKEMPI1 ($\rho = 0.57$ and RMSE = 1.57 kcal/mol).

In addition, we evaluated the performance of our approach on a subset of 472 mutations (S472) not present within the first version of SKEMPI but included in SKEMPI2. For this experiment, we trained a predictive model using all variants from the first version of SKEMPI (S1964). Our method achieved a correlation of 0.63 (RMSE = 1.11 kcal/mol).

Validation on CAPRI

mCSM-PPI2 was further validated against the CAPRI (52) round 26, which is composed of 1862 experimentally characterised mutations in two *de novo* influenza inhibitor targets (T55 and T56: 1007 mutations at 53 different positions

Table 1. Comparative performance of mCSM-PPI2 on CAPRI and the blind test set for the complex MDM2-P53

	CAPRI (T55)		CAPRI (T56)		MDM2-P53	
Method	Kendall	RMSE (kcal/mol)	Kendall	RMSE (kcal/mol)	ρ	RMSE (kcal/mol)
mCSM-PPI2	**0.42**	**2.55**	**0.32**	**4.06**	**0.40**	**0.36**
mCSM (20)	0.16**	3.71	0.13**	4.15	0.23	0.83
MutaBind (25)	0.41	2.58	0.30	4.27	*NA*	*NA*
iSEE (24)	*NA*	*NA*	*NA*	*NA*	0.24	0.81
BeAtMuSiC (23)	0.28**	3.04	0.30	4.06	−0.23*	0.91
FoldX (22)	0.12**	3.94 0.90	0.16**	4.33	−0.14*	
MMPBSA (21)	0.19**	5.40	0.08**	28.04	*NA*	*NA*

*p-value < 0.05 by Fisher r-to-z transformation test compared to mCSM-PPI2
**p-value < 0.05 by transforming tau-to-r followed by Fisher r-to-z transformation. *NA*: Data not available.

in T55 and 855 mutations at 45 different positions in T56). The *in vitro* experimental measurements used the enrichment values generated from deep sequencing and were calculated based on the binary logarithm of the ratio of number of times the variant sequence was observed after and before the selection for binding. Although the 3D structures for these two complexes were not available, structures of close homologues have been described (53,54) and were used for generating homology model structures by introducing point mutations using Modeller (55) (see Supplementary Materials). mCSM-PPI2 was able to achieve a Kendall's score of up to 0.42 and 0.32 for mutations in T55 and T56, respectively, ranking first amongst 26 other methods (Supplementary Figure S10 and Table 1).

Blind test

The performance of mCSM-PPI2 was further evaluated on a small set of 26 variants at the interface of interaction of the MDM2-p53 complex (PDB 1YCR) (24). Our method achieved a Pearson's Correlation of 0.40 and an RMSE = 0.36 kcal/mol outperforming mCSM, iSEE, FoldX, BeAtMuSiC (23) (Table 1).

Finally, we looked at the ability of mCSM-PPI2 to accurately identify PPI hotspots, residues that contribute to the majority of the binding free energy of the interaction and have been recognized as important sites for drug development (5). Here we evaluated the performance of mCSM-PPI2 across a previously proposed set of 378 alanine-scanning experimental mutations within 19 different protein–protein complexes (56,57) (Supplementary Table S7). In order to minimize biases, for this experiment we removed 232 variants from S8338 which were redundant with our set of 378 alanine scanning mutations. Our predictive model was was able to accurately distinguish hot and not-hot spots (95% of hotspots and 92% of non-hotspots were correctly predicted) outperforming the results reported for Robetta (precision of 79% and 68% when predicting hotspots and non-hotspots, respectively). The predicted changes in binding energy showed that mCSM-PPI2 predictions correlated strongly with the experimental data (Pearson's Correlation of 0.95 and RMSE of 0.25 kcal/mol; Supplementary Figure S11). These results indicate that mCSM-PPI2 could also be a powerful tool for hotspot identification.

CONCLUSION

Here, we introduce mCSM-PPI2, a web server that implements an integrated computation approach for predicting effects of missense mutations in protein–protein affinity. By consolidating our graph-based signatures framework with evolutionary information, inter-atomic contacts and energy terms our updated method has shown to perform better than its previous version and other methods. In addition, the use of hypothetical reverse mutations has shown to improve the robustness of our predictive model allowing for a more balanced prediction. mCSM-PPI2 is freely available as a user-friendly and easy to use web server at http://biosig.unimelb.edu.au/mcsm_ppi2/.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Australian Government Research Training Program Scholarship [to C.H.M.R and Y.M.]; Jack Brockhoff Foundation [JBF 4186, 2016 to D.B.A.]; a Newton Fund RCUK-CONFAP Grant awarded by The Medical Research Council (MRC) and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) [MR/M026302/1 to D.B.A. and D.E.V.P.]; National Health and Medical Research Council of Australia [APP1072476 to D.B.A.]; Victorian Life Sciences Computation Initiative (VLSCI), an initiative of the Victorian Government, Australia, on its Facility hosted at the University of Melbourne [UOM0017]; Instituto René Rachou (IRR/FIOCRUZ Minas), Brazil and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [to D.E.V.P.]; Department of Biochemistry and Molecular Biology, University of Melbourne [to D.B.A.]; Supported in part by the Victorian Government's OIS Program. Funding for open access charge: MRC.
Conflict of interest statement. None declared.

REFERENCES

1 Gao,J., Li,W.X., Feng,S.Q., Yuan,Y.S., Wan,D.F., Han,W. and Yu,Y. (2008) A protein–protein interaction network of transcription factors acting during liver cell proliferation. *Genomics*, **91**, 347–355.

2 Chuderland,D. and Seger,R. (2005) Protein-protein interactions in the regulation of the extracellular signal-regulated kinase. *Mol. Biotechnol.*, **29**, 57–74.
3 Nicod,C., Banaei-Esfahani,A. and Collins,B.C. (2017) Elucidation of host-pathogen protein–protein interactions to uncover mechanisms of host cell rewiring. *Curr. Opin. Microbiol.*, **39**, 7–15.
4 Paumi,C.M., Menendez,J., Arnoldo,A., Engels,K., Iyer,K.R., Thaminy,S., Georgiev,O., Barral,Y., Michaelis,S. and Stagljar,I. (2007) Mapping protein–protein interactions for the yeast ABC transporter Ycf1p by integrated split-ubiquitin membrane yeast two-hybrid analysis. *Mol. Cell*, **26**, 15–25.
5 Jubb,H., Blundell,T.L. and Ascher,D.B. (2015) Flexibility and small pockets at protein–protein interfaces: New insights into druggability. *Prog. Biophys. Mol. Biol.*, **119**, 2–9.
6 Gao,M., Zhou,H and Skolnick,J. (2015) Insights into disease-associated mutations in the human proteome through protein structural analysis. *Structure*, **23**, 1362–1369.
7 David,A., Razali,R., Wass,M.N. and Sternberg,M.J. (2012) Protein-protein interaction sites are hot spots for disease-associated nonsynonymous SNPs. *Hum. Mutat.*, **33**, 359–363.
8 Engin,H.B., Kreisberg,J.F. and Carter,H. (2016) Structure-based analysis reveals cancer missense mutations target protein interaction interfaces. *PLoS One*, **11**, e0152929.
9 Jubb,H.C., Pandurangan,A.P., Turner,M.A., Ochoa-Montano,B., Blundell,T.L. and Ascher,D.B. (2017) Mutations at protein–protein interfaces: small changes over big surfaces have large impacts on human health. *Prog. Biophys. Mol. Biol.*, **128**, 3–13.
10 Ascher,D.B., Spiga,O., Sekelska,M., Pires,D.E.V., Bernini,A., Tiezzi,M., Kralovicova,J., Borovska,I., Soltysova,A., Olsson,B. *et al.* (2019) Homogentisate 1,2-dioxygenase (HGD) gene variants, their analysis and genotype-phenotype correlations in the largest cohort of patients with AKU. *Eur. J. Hum. Genet.*, **27**, 888–902.
11 Hnizda,A., Fabry,M., Moriyama,T., Pachl,P., Kugler,M., Brinsa,V., Ascher,D.B., Carroll,W.L., Novak,P., Zaliova,M. *et al.* (2018) Relapsed acute lymphoblastic leukemia-specific mutations in NT5C2 cluster into hotspots driving intersubunit stimulation. *Leukemia*, **32**, 1393–1403.
12 Andrews,K.A., Ascher,D.B., Pires,D.E.V., Barnes,D.R., Vialard,L., Casey,R.T., Bradshaw,N., Adlard,J., Aylwin,S., Brennan,P. *et al.* (2018) Tumour risks and genotype-phenotype correlations associated with germline variants in succinate dehydrogenase subunit genes SDHB, SDHC and SDHD. *J. Med. Genet.*, **55**, 384–394.
13 Soardi,F.C., Machado-Silva,A., Linhares,N.D., Zheng,G., Qu,Q., Pena,H.B., Martins,T.M.M., Vieira,H.G.S., Pereira,N.B., Melo-Minardi,R.C. *et al.* (2017) Familial STAG2 germline mutation defines a new human cohesinopathy. *NPJ Genom. Med.*, **2**, 7.
14 Nemethova,M., Radvanszky,J., Kadasi,L., Ascher,D.B., Pires,D.E., Blundell,T.L., Porfirio,B., Mannoni,A., Santucci,A., Milucci,L. *et al.* (2016) Twelve novel HGD gene variants identified in 99 alkaptonuria patients: focus on 'black bone disease' in Italy. *Eur. J. Hum. Genet.*, **24**, 66–72.
15 Jafri,M., Wake,N.C., Ascher,D.B., Pires,D.E., Gentle,D., Morris,M.R., Rattenberry,E., Simpson,M.A., Trembath,R.C., Weber,A. *et al.* (2015) Germline mutations in the CDKN2B tumor suppressor gene predispose to renal cell carcinoma. *Cancer Discov.*, **5**, 723–729.
16 Blaszczyk,M., Harmer,N.J., Chirgadze,D.Y., Ascher,D.B. and Blundell,T.L. (2015) Achieving high signal-to-noise in cell regulatory systems: Spatial organization of multiprotein transmembrane assemblies of FGFR and MET receptors. *Prog. Biophys. Mol. Biol.*, **118**, 103–111.
17 Ascher,D.B., Wielens,J., Nero,T.L., Doughty,L., Morton,C.J. and Parker,M.W. (2014) Potent hepatitis C inhibitors bind directly to NS5A and reduce its affinity for RNA. *Sci. Rep.*, **4**, 4765.
18 Portelli,S., Phelan,J.E., Ascher,D.B., Clark,T.G. and Furnham,N. (2018) Understanding molecular consequences of putative drug resistant mutations in Mycobacterium tuberculosis. *Sci. Rep.*, **8**, 15356.
19 Vedithi,S.C., Malhotra,S., Das,M., Daniel,S., Kishore,N., George,A., Arumugam,S., Rajan,L., Ebenezer,M., Ascher,D.B. *et al.* (2018) Structural implications of mutations conferring rifampin resistance in mycobacterium leprae. *Sci. Rep.*, **8**, 5016.
20 Pires,D.E., Ascher,D.B. and Blundell,T.L. (2014) mCSM: Predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics*, **30**, 335–342.
21 Kollman,P.A., Massova,I., Reyes,C., Kuhn,B., Huo,S., Chong,L., Lee,M., Lee,T., Duan,Y., Wang,W. *et al.* (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.*, **33**, 889–897.
22 Schymkowitz,J., Borg,J., Stricher,F., Nys,R., Rousseau,F. and Serrano,L. (2005) The FoldX web server: an online force field. *Nucleic Acids Res.*, **33**, W382–W388.
23 Dehouck,Y., Kwasigroch,J.M., Rooman,M. and Gilis,D. (2013) BeAtMuSiC: Prediction of changes in protein–protein binding affinity on mutations. *Nucleic Acids Res.*, **41**, W333–W339.
24 Geng,C., Vangone,A., Folkers,G.E., Xue,L.C. and Bonvin,A. (2019) iSEE: Interface structure, evolution, and energy-based machine learning predictor of binding affinity changes upon mutations. *Proteins*, **87**, 110–119.
25 Li,M., Simonetti,F.L., Goncearenco,A. and Panchenko,A.R. (2016) MutaBind estimates and interprets the effects of sequence variants on protein–protein interactions. *Nucleic Acids Res.*, **44**, W494–W501.
26 Pires,D.E., Ascher,D.B. and Blundell,T.L. (2014) DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Res.*, **42**, W314–W319.
27 Rodrigues,C.H., Pires,D.E. and Ascher,D.B. (2018) DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Res.*, **46**, W350–W355.
28 Rodrigues,C.H., Ascher,D.B. and Pires,D.E. (2018) Kinact: a computational approach for predicting activating missense mutations in protein kinases. *Nucleic Acids Res.*, **46**, W127–W132.
29 Pires,D.E.V. and Ascher,D.B. (2017) mCSM-NA: predicting the effects of mutations on protein-nucleic acids interactions. *Nucleic Acids Res.*, **45**, W241–W246.
30 Pires,D.E., Chen,J., Blundell,T.L. and Ascher,D.B. (2016) In silico functional dissection of saturation mutagenesis: Interpreting the relationship between phenotypes and changes in protein stability, interactions and activity. *Sci. Rep.*, **6**, 19848.
31 Pires,D.E., Blundell,T.L. and Ascher,D.B. (2016) mCSM-lig: quantifying the effects of mutations on protein-small molecule affinity in genetic disease and emergence of drug resistance. *Sci. Rep.*, **6**, 29575.
32 Pires,D.E. and Ascher,D.B. (2016) CSM-lig: a web server for assessing and comparing protein-small molecule affinities. *Nucleic Acids Res.*, **44**, W557–W561.
33 Pires,D.E. and Ascher,D.B. (2016) mCSM-AB: a web server for predicting antibody-antigen affinity changes upon mutation with graph-based signatures. *Nucleic Acids Res.*, **44**, W469–W473.
34 Pires,D.E., Blundell,T.L. and Ascher,D.B. (2015) Platinum: a database of experimentally measured effects of mutations on structurally defined protein-ligand complexes. *Nucleic Acids Res.*, **43**, D387–D391.
35 Moal,I.H. and Fernandez-Recio,J. (2012) SKEMPI: a structural kinetic and energetic database of mutant protein interactions and its use in empirical models. *Bioinformatics*, **28**, 2600–2607.
36 Jankauskaite,J., Jimenez-Garcia,B., Dapkunas,J., Fernandez-Recio,J. and Moal,I.H. (2019) SKEMPI 2.0: an updated benchmark of changes in protein–protein binding energy, kinetics and thermodynamics upon mutation. *Bioinformatics*, **35**, 462–469.
37 Sirin,S., Apgar,J.R., Bennett,E.M. and Keating,A.E. (2016) AB-Bind: antibody binding mutational database for computational affinity predictions. *Protein Sci.*, **25**, 393–409.
38 Jemimah,S., Yugandhar,K. and Michael Gromiha,M. (2017) PROXiMATE: a database of mutant protein–protein complex thermodynamics and kinetics. *Bioinformatics*, **33**, 2787–2788.
39 Liu,Q., Chen,P., Wang,B., Zhang,J. and Li,J. (2018) dbMPIKT: a database of kinetic and thermodynamic mutant protein interactions. *BMC Bioinformatics*, **19**, 455.
40 Thiltgen,G. and Goldstein,R.A. (2012) Assessing predictors of changes in protein stability upon mutation using self-consistency. *PLoS One*, **7**, e46084.
41 Petukh,M., Dai,L. and Alexov,E. (2016) SAAMBE: Webserver to predict the charge of binding free energy caused by amino acids mutations. *Int. J. Mol. Sci.*, **17**, 547.
42 Pires,D.E., Blundell,T.L. and Ascher,D.B. (2015) pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J. Med. Chem.*, **58**, 4066–4072.

43 Sahni,N., Yi,S., Taipale,M., Fuxman Bass,J.I., Coulombe-Huntington,J., Yang,F., Peng,J., Weile,J., Karras,G.I., Wang,Y. *et al.* (2015) Widespread macromolecular interaction perturbations in human genetic disorders. *Cell*, **161**, 647–660.

44 Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.

45 Cock,P.J., Antao,T., Chang,J.T., Chapman,B.A., Cox,C.J., Dalke,A., Friedberg,I., Hamelryck,T., Kauff,F., Wilczynski,B. *et al.* (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, **25**, 1422–1423.

46 Chen,Z., Zhao,P., Li,F., Leier,A., Marquez-Lago,T.T., Wang,Y., Webb,G.I., Smith,A.I., Daly,R.J., Chou,K.C. *et al.* (2018) iFeature: a Python package and web server for features extraction and selection from protein and peptide sequences. *Bioinformatics*, **34**, 2499–2502.

47 Jubb,H.C., Higueruelo,A.P., Ochoa-Montano,B., Pitt,W.R., Ascher,D.B. and Blundell,T.L. (2017) Arpeggio: a web server for calculating and visualising interatomic interactions in protein structures. *J. Mol. Biol.*, **429**, 365–371.

48 Li,G., Semerci,M., Yener,B. and Zaki,M.J. (2012) Effective graph classification based on topological and label attributes. *Stat. Anal. Data Mining: ASA Data Sci. J.*, **5**, 265–283.

49 Kawashima,S. and Kanehisa,M. (2000) AAindex: amino acid index database. *Nucleic Acids Res.*, **28**, 374.

50 Grant,B.J., Rodrigues,A.P., ElSawy,K.M., McCammon,J.A. and Caves,L.S. (2006) Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics*, **22**, 2695–2696.

51 Rose,A.S., Bradley,A.R., Valasatava,Y., Duarte,J.M., Prlic,A. and Rose,P.W. (2018) NGL viewer: web-based molecular graphics for large complexes. *Bioinformatics*, **34**, 3755–3758.

52 Janin,J., Henrick,K., Moult,J., Eyck,L.T., Sternberg,M.J., Vajda,S., Vakser,I., Wodak,S.J. and Critical Assessment of, P.I. (2003) CAPRI: a critical assessment of predicted interactions. *Proteins*, **52**, 2–9.

53 Fleishman,S.J., Whitehead,T.A., Ekiert,D.C., Dreyfus,C., Corn,J.E., Strauch,E.M., Wilson,I.A. and Baker,D. (2011) Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science*, **332**, 816–821.

54 Whitehead,T.A., Chevalier,A., Song,Y., Dreyfus,C., Fleishman,S.J., De Mattos,C., Myers,C.A., Kamisetty,H., Blair,P., Wilson,I.A. *et al.* (2012) Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat. Biotechnol.*, **30**, 543–548.

55 Sali,A. and Blundell,T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.

56 Ascher,D.B., Jubb,H.C., Pires,D.E., Ochi,T., Higueruelo,A. and Blundell,T.L. (2015) Protein-protein interactions: structures and druggability. In: Scapin,G, Patel,D and Arnold,E (eds). *Multifaceted Roles of Crystallography in Modern Drug Discovery*. Springer, Netherlands, pp. 141–163.

57 Kortemme,T. and Baker,D. (2002) A simple physical model for binding energy hot spots in protein–protein complexes. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 14116–14121.